

Tech Review

(digital library information retrieval model)

Jiacheng Guo(jg17)

1. Introduction

As the carrier of knowledge economy, digital library has become the focus in the competition of culture, science, and technology. The emergence of big data analysis technology has changed the traditional library, which is limited by region, time, and space, and can provide useful information resources for people at any time.

2. Body

The traditional retrieval model is based on keyword information retrieval, which has been unable to meet the needs of readers. This is caused by the low level of intelligence and sharing of information retrieval query services. In order to solve this problem, on the basis of the original, retrieval information from the knowledge level, understand the user's query request from the semantic point of view, and analyze the information retrieval process. The information retrieval model of Digital Library Based on big data analysis technology deals with the retrieval needs by analyzing the information set that meets the needs of users.

Using big data analysis technology to build domain ontology, and then to the knowledge base to search the information users need to retrieve. When the user inputs the query request, the reasoning opportunity infers the keywords or statements based on the domain ontology to form a logical expression and upload it to the retrieval system. RDF and other tools should be used to describe case-based reasoning. According to reasoning rules, system programmers create rules according to specific conditions. Finally, the relevant literature resources that meet the needs of users are searched out from the information resources of digital library.

However, due to the different types of domain ontologies, domain ontologies need to be constructed according to the level of ontology development. The specific construction methods are as follows:

- 1) The single ontology method is used to explain the shared vocabulary set provided by the global ontology. All information sources must relate to the global ontology to ensure semantic consistency. When a specific domain needs to be mapped, this method can be used, provided that the single ontology is not affected by the change of information sources

2) The multi ontology method is used to describe different information sources, and each ontology has its own vocabulary set. When the information source changes, the ontology structure changes little, which is the advantage of multi ontology method. According to the requirements of domain ontology, the appropriate classification method is selected to accurately search the relevant literature that meets the user needs from the information resources of digital library.

After the first mock exam, users need to process user query information and use unified mode to parse text form and store it to the document database. Under the Marc standard, the document information in the document database is extracted according to the standard definition of the number of digital document elements. In order to facilitate data sharing, XML is used to extract metadata and store it in metadata database.

According to Marc metadata extraction standard, various database information is extracted and refined to realize document information metadata extraction. However, because XML does not have the ability of semantic description, it is necessary to build a conceptual model to describe the relationship between them.

After completing the above work, we use semantic dictionary and corpus vocabulary to simplify the program and reduce the workload of creating metadata.

The model contains four levels. The first layer which is also called the application layer, is based on the Internet portal set, combined with the characteristics of artificial intelligence services, is convenient for users to access information in various functional interfaces. The second layer: service layer, the core part of the model. It is mainly responsible for intelligent information retrieval service and personalized information push service. The third layer: support layer is the engine part of digital library information retrieval platform. It is mainly responsible for collecting information, retrieving business, classifying information, and storing information. The fourth layer: resource layer is the support of the whole digital library's own database.

The model is based on the principle of conceptual semantic retrieval, and the specific retrieval process is as follows: Firstly, using concept retrieval technology, with the help of domain experts, construct domain ontology; secondly, in Marc In the end, it uses the similarity degree of big domain ontology to search, submit to the large domain metadata, and then submit it to the meta data base. Thus, the design of digital library information retrieval model based on big data analysis technology is completed.

3. Conclusion

The digital library information retrieval model of big data analysis technology provides a sharing platform. Compared with the traditional model, big data analysis technology has broad application prospects and can build personalized information

recommendation and intelligent information retrieval model.

Reference:

1. <https://science.sciencemag.org/content/275/5298/327>
2. https://www.ercim.eu/publication/Ercim_News/enw27/ferber.html
3. http://disi.unitn.it/~bernardi/Courses/DL/Slides_11_12/4.pdf